

CLAIMS:

- 1 1. A method for determining patterns in an input string of tokens, comprising
2 steps of:
3 identifying extensible patterns in the input string;
4 creating an inexact tree for the input string, using the patterns identified; and
5 displaying a set of extensible patterns identified by the inexact tree;
6 wherein creating the inexact tree comprises creating nodes and edges,
7 connecting the nodes,
8 wherein each node represents a subset of a string and each edge connects a
9 lower order node to a higher order node;
10 wherein each subset comprises a pattern comprising extensible string; and
11 wherein each extensible string comprises at least one dot token.
- 1 2. The method of claim 1, further comprising receiving a parameter k specifying
2 the minimum times an extensible pattern must occur in a sequence.
- 1 3. The method of claim 1, wherein the step of identifying patterns in the input
2 string B comprises creating a rigid string m' from an extensible string m .
- 1 4. The method of claim 1 wherein the step of identifying patterns in the input
2 string B comprises extracting a subset of tokens b from the input string B .
- 1 5. The method of claim 4 analyzing the subset of tokens b to determine whether
2 the subset is compatible with the rigid string m' .

1 6. The method of claim 5 wherein if the subset b is compatible with the rigid
2 string m' the subset and the rigid string are concatenated into a new rigid string m_t .

1 7. The method of claim 6 further comprising the step of running a routine for
2 determining whether the concatenated string is non maximal with respect to its nodes
3 of the same order.

1 8. The method of claim 7 further comprising removing each node form the tree
2 that is non maximal with respect to its nodes of the same order.

1 9. The method of claim 8 wherein if the magnitude of the location list of the rigid
2 string m' is equal to the magnitude of the location list of the subset of tokens b then
3 the size of the collection of tokens B is reduced by removing the subset of tokens b
4 determined in the step of extracting a subset of tokens from the input string.

1 10. The method of claim 9 wherein if the number of times the rigid string pattern
2 repeats is greater than the minimum number of times an extensible pattern must occur
3 in a sequence k , then the concatenated extensible string m_t is converted into a rigid
4 string m' .

- 1 11. The method of claim 10 wherein the method of claim 1 is performed on the
2 converted rigid string m' .
- 1 12. The method of claim 11 further comprising identifying a zone for each
2 subsequence of tokens Z_r such that each occurrence of each pattern is fully contained
3 within the zone of the rigid string $Z_{m'}$.
- 1 13. The method of claim 11 further comprising determining whether the rigid
2 string m' is not maximal with respect to a string of tokens r that are returned from the
3 determination of the routine.
- 1 14. The method of claim 13 wherein the result of the routine m' is added to a
2 collection of maximal extensible patterns *Result*.

1 15. A system comprising:
2 an input/output device for receiving information including an input string; and
3 a processor for identifying extensible patterns; and
4 a memory for storing identified patterns and for storing the inexact suffix tree.

1 16. The system of claim 15 wherein the input/output device further comprising a
2 CD ROM drive.

1 17. The system of claim 15 wherein the input/output device further comprises a
2 network interface.

1 18. The system of claim 15 wherein the memory further comprises an operating
2 system.

1 19. The system of claim 15 wherein the memory further comprises an application.

1 20. A program product for determining patterns in an input string of tokens,
2 comprising instructions for:
3 identifying extensible patterns in the input string;
4 creating an inexact tree for the input string, using the patterns identified; and
5 displaying a set of extensible patterns identified by the inexact tree;
6 wherein creating the inexact tree comprises creating nodes and edges,
7 connecting the nodes,
8 wherein each node represents a subset of a string and each edge connects a
9 lower order node to a higher order node;
10 wherein each subset comprises a pattern comprising extensible string; and
11 wherein each extensible string comprises at least one dot token.